Міністерство освіти і науки України

Львівський національний університет імені Івана Франка

Факультет електроніки та комп’ютерних технологій

Кафедра системного проектування

Звіт

Про виконання лабораторної роботи №4

З курсу «Основи аналізу даних»

### **Кореляційний аналіз**

Виконав:

Студент групи ФЕІ-42

Рейнгардт В.В.

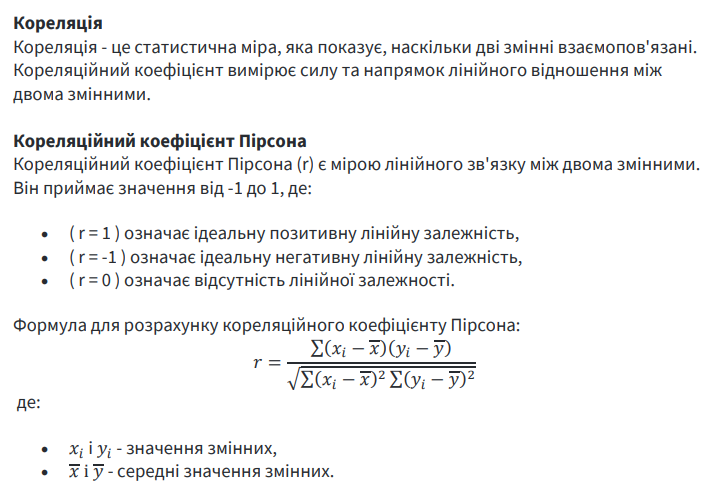
Перевірив:

Асист. Азаров І.В.

Львів 2024

**Мета:** Ознайомитися з методами аналізу кореляції між змінними в наборах даних, навчитися візуалізувати кореляційні матриці та інтерпретувати кореляційні коефіцієнти.

**Теоретичні відомості:**



**Хід роботи:**

1. Підготовка даних
   1. Завантажте набір даних про діабет.

*import* pandas *as* pd

*# Завантаження набору даних*

url ='https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv'

column\_names = ['Pregnancies', 'Glucose', 'BloodPressure',

'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction',

'Age', 'Outcome']

diabetes\_df = pd.read\_csv(url, names=column\_names)

*# Відображення перших кількох рядків*

print(diabetes\_df.head())

* 1. Виконайте попередню обробку даних, включаючи заповнення пропущених значень та видалення непотрібних змінних.

*# Перевірка на наявність пропущених значень*

print(diabetes\_df.isnull().sum())

*# Приклад: якщо є пропущені значення, заповнимо середніми значеннями*

df = diabetes\_df.fillna(diabetes\_df.mean())

1. Розрахунок кореляції
   1. Розрахуйте кореляційну матрицю для всіх змінних в наборі даних.

correlation\_matrix = diabetes\_df.corr()

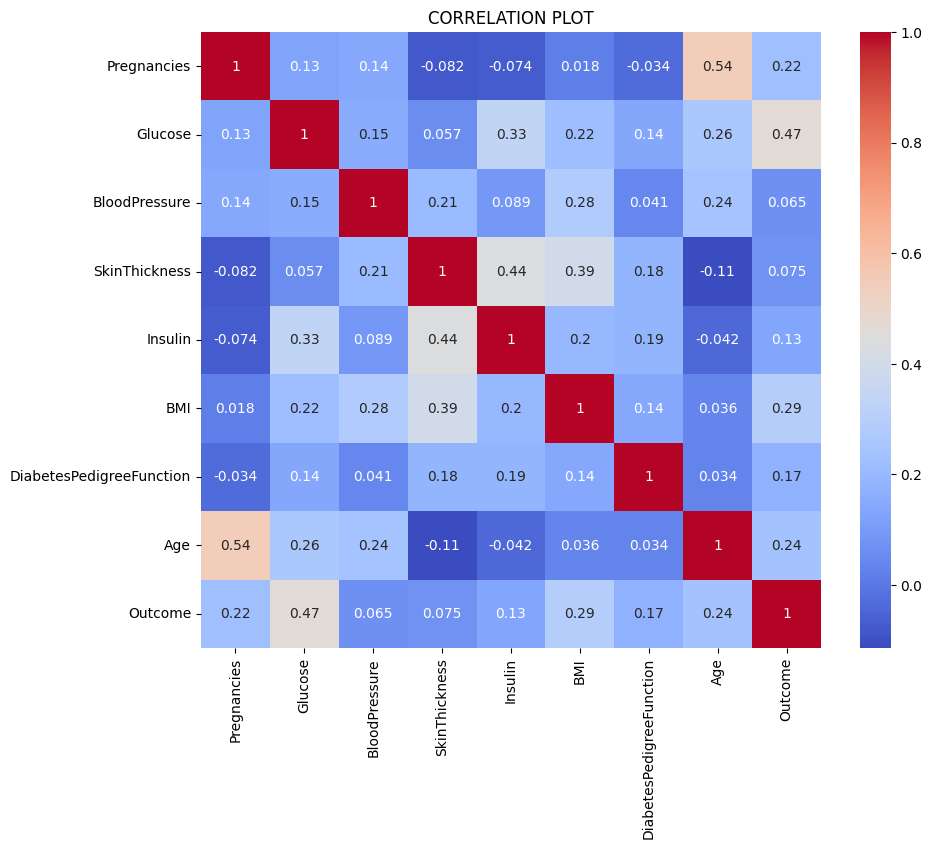
2.2 Використовуйте бібліотеку Seaborn для побудови теплової карти кореляційної матриці.

plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix,annot=True,cmap='coolwarm')

plt.title('CorreLATION Plot')

plt.show



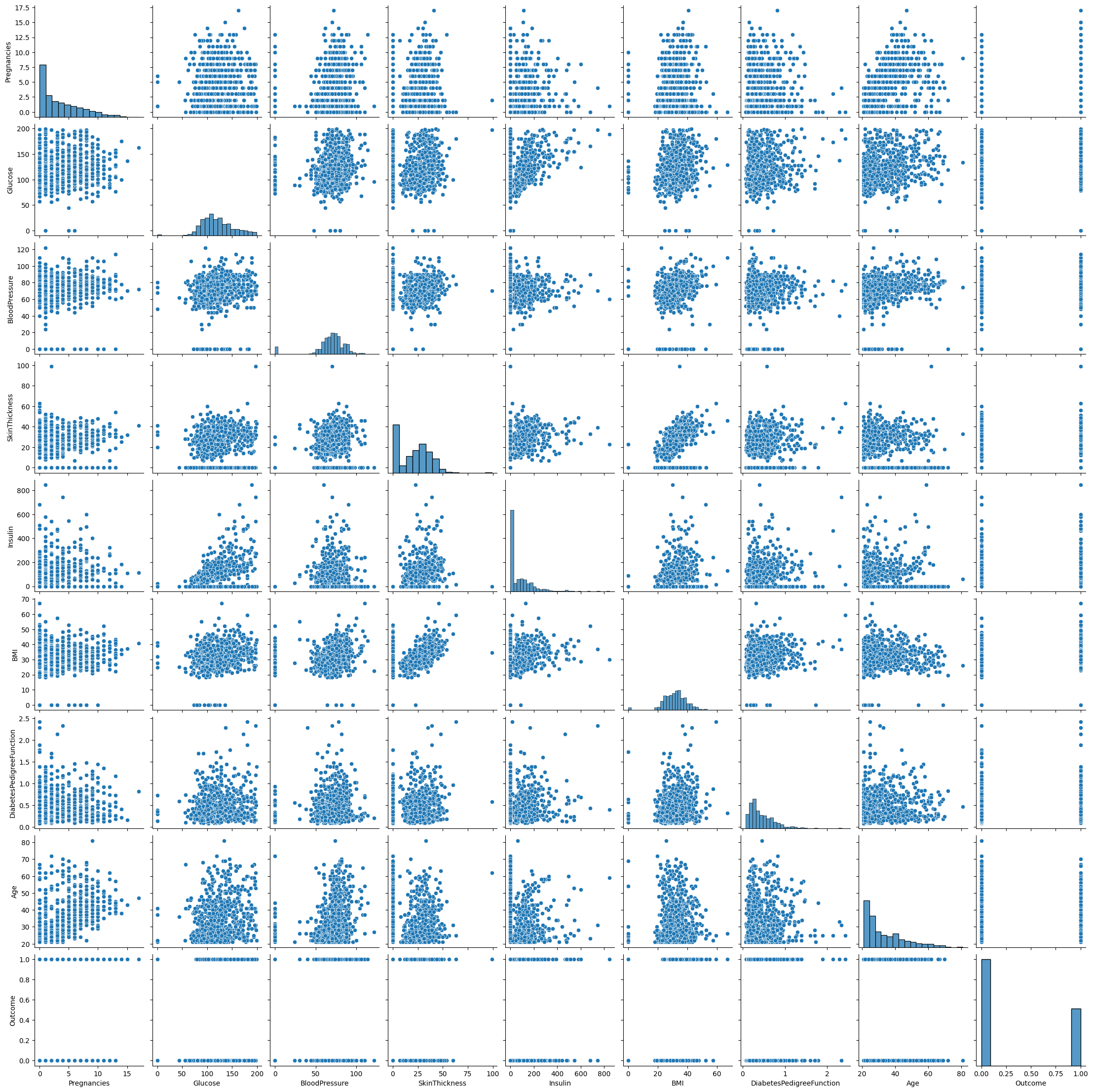
1. Візуалізація кореляції
   1. Використовуйте функцію pairplot() з бібліотеки Seaborn для візуалізації парних графіків між змінними.

*# Використовуємо pairplot для візуалізації парних графіків*

sns.pairplot(diabetes\_df)

*# Показуємо графік*

plt.show()



1. Інтерпретація результатів
   1. Проаналізуйте кореляційну матрицю та виявите змінні, які мають сильну позитивну або негативну кореляцію.

З кореляційної матриці видно, що жодні із змінних не мають сильну позитивну або негативну кореляцію. Хіба можна виділити середню кореляцію між Віком і Кількістю вагітностей, а також між Товщиною шкіри і Рівнем інсуліну, а також, між Рівнем глюкози і Наявністю діабету.

* 1. Інтерпретуйте парні графіки для виявлення можливих лінійних або нелінійних залежностей між змінними.

З парних графіків можна виділити лінійну залежність між BMI і Товщиною шкіри. Решта змінні не мають явних лінійних чи нелінійних залежностей.

**Висновок:**

У ході виконання лабораторної роботи було проведено кореляційний аналіз на основі набору даних про діабет. Ми ознайомилися з методами виявлення зв'язків між змінними та оцінили їх інтенсивність за допомогою кореляційної матриці. Результати візуалізовано у вигляді теплової карти та парних графіків, що дало змогу краще зрозуміти взаємозалежності між показниками.